



## Comparison of different methods of incorporating respiratory motion for lung cancer tumor volume delineation on PET images: a simulation study.

Mathieu Hatt, Amandine Le Maitre, Daphné Wallach, Hadi Fayad, Dimitris Visvikis

### ► To cite this version:

Mathieu Hatt, Amandine Le Maitre, Daphné Wallach, Hadi Fayad, Dimitris Visvikis. Comparison of different methods of incorporating respiratory motion for lung cancer tumor volume delineation on PET images: a simulation study.. *Physics in Medicine and Biology*, 2012, 57 (22), pp.7409-30. 10.1088/0031-9155/57/22/7409 . hal-00749010

**HAL Id: hal-00749010**

**<https://hal.science/hal-00749010>**

Submitted on 6 Nov 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Comparison of different methods of incorporating respiratory motion for lung cancer tumor volume delineation on PET images: a simulation study**

Incorporation of motion information for lung tumor volume definition in PET

Mathieu Hatt<sup>@\*</sup>, Amandine Le Maitre<sup>1\*</sup>, Daphné Wallach<sup>2\*</sup>, Hadi Fayad, and Dimitris Visvikis

<sup>\*</sup>First three authors listed in alphabetical order contributed equally to this work

INSERM, UMR 1101 LaTIM, CHRU Morvan, Brest F-29200, France

@ corresponding author

Mathieu Hatt

INSERM UMR 1101 LaTIM

CHRU Morvan

Brest, 29609 CEDEX

France

hatt@univ-brest.fr

tel: +33 2 98 01 81 11

Wordcount: >7986

---

<sup>1</sup> Amandine Le Maitre is currently with the LTSI INSERM UMR 1099, University of Rennes 1, Rennes, France

<sup>2</sup> Daphné Wallach is currently with the ARTORG Center for Computer Aided Surgery, University of Bern, Bern, Switzerland.

**Abstract.** The interest of PET complementary information for the delineation of the target volume in radiotherapy of lung cancer is increasing. However, respiratory motion requires the determination of a functional internal target volume (ITV) on PET images for which several strategies have been proposed. The purpose of this study was the comparison of these strategies for taking into account respiratory motion and deriving the ITV: 1) adding fixed margins to the volume defined on a single binned image, 2) segmenting a motion averaged image, and 3) considering the union of volumes delineated on binned frames. For this third strategy, binned frames were either non-corrected for motion, or corrected using two different methods: elastic registration or super resolution. The strategies' performances were assessed on realistic simulated datasets combining the NCAT phantom with a PET Philips GEMINI scanner model in GATE, and containing various configurations of tumor to background contrast, with both regular and irregular respiratory motion (with a range of motion amplitudes). The obtained ITVs' sensitivity (SE) and positive predictive value (PVE) with respect to the known true ITV were significantly higher (from 0.8 to 0.95) than all other techniques when using binned frames corrected for motion, independently of motion regularity, amplitude, or tumor to background contrast. Although the absolute difference was small and not always significant, images corrected using super resolution led to systematically better results than using elastic registration. The worst results were obtained when using the motion averaged image for sensitivity (around 0.5-0.6) and using the margins added to a single frame for PPV (0.6-0.7) respectively. The best strategy to account for breathing motion for tumor ITV delineation in radiotherapy planning is to rely on the use of the union of volumes delineated on super resolution-corrected binned images.

**Keywords:** radiotherapy treatment planning, respiratory motion, super-resolution, 4D PET imaging

## 1. Introduction

Computed tomography (CT) is routinely used for delineating gross target volumes (GTV) and organs at risk (OAR) in radiotherapy treatment planning (RTP) of lung cancer. However, the high sensitivity and specificity of  $^{18}\text{F}$ -FDG Positron Emission Tomography (PET) for diagnosis and staging, as well as the widespread availability of combined PET/CT scanners, led to numerous studies investigating the potential of PET/CT for target volume delineation. Use of PET/CT images has been shown to reduce intra and inter observer variability in target volume delineation [Ashamalla 2005, Steenbakkers 2006, van Baardwijk 2007], as well as significantly modifying GTVs [Walter 1999, Bradley 2004]. Moreover, PET imaging allows characterizing tracer uptake heterogeneities than can be exploited for the prescription of inhomogeneous doses in scenarios of dose boosting, painting or redistribution [Aristophanous 2011, van Loon 2011]. However, metabolically active tumor volume (MATV) delineation faces a number of challenges due to the poor spatial resolution and high level of noise characterizing emission images [Hatt 2011a].

Respiratory motion constitutes an additional challenge for lung cancer RTP. In order to thoroughly irradiate the tumor while sparing OARs, it is desirable to define the smallest target volume encompassing the tumor in each phase of the respiratory cycle. Visualizing the tumor morphological volume and location throughout the respiratory cycle, requires a four-dimensional (4D) CT scan. The overall internal target volume (ITV) is then defined as the union of all GTV volumes delineated on the images corresponding to all phases of the respiratory cycle [De Bari 2011].

On the other hand, the functional ITV definition may be performed on a standard PET image under the assumption that a motion averaged image represents the tumor motion extent [Riegel 2012, Caldwell 2003], or through the addition of fixed margins on a single respiration synchronized frame image [De Bari 2011]. Alternatively, 4D imaging can be also used to define a functional ITV on PET images [Nehmeh 2004, Wolthaus 2005, Aristophanous 2012], which requires the MATV delineation on each respiratory synchronized PET frame. The main advantage of binned frames over the motion average image is that they should in principle more accurately reflect the total extent of the MATV motion displacement. However, one significant disadvantage with respect to the motion averaged image is the reduced quality of the individual respiratory synchronized image frames [Boucher 2004,

Jarritt 2006, Park 2008], since each temporal frame contains only a fraction of the overall acquired data. Consequently, the reconstructed respiratory synchronized PET images are characterized by lower signal to noise ratio and contrast [Jarritt 2006]. Such a decreased image quality should in turn lead to increased errors in overall MATV determination considering the current clinically available PET segmentation algorithms [Hatt 2011b]. On the other hand, several methods allowing the correction of motion artifacts and the improvement of statistical quality of respiratory synchronized PET images have been proposed. These include both ‘post-reconstruction’ or ‘reconstruction’ based approaches, making use of affine [Lamare 2007a] and/or deformable respiratory motion models [Qiao 2006, Lamare 2007b] or making use of the superresolution principle [Chang 2009, Wallach 2012].

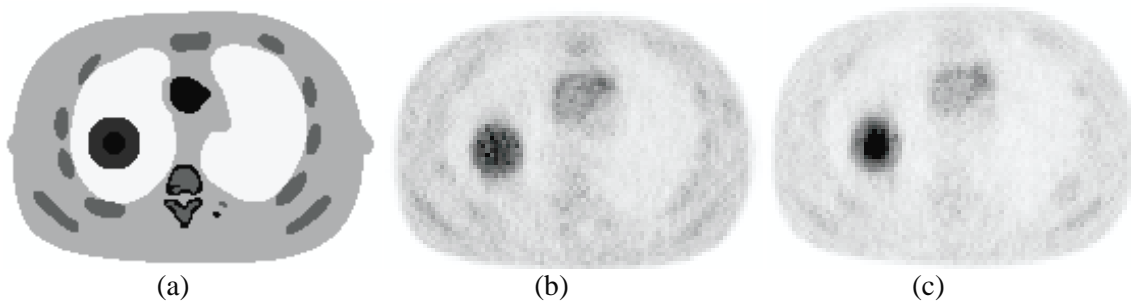
All of these different ways of taking into account the respiratory motion can have a significant influence on the determination of the functional GTV/ITV using PET images. To the best of our knowledge, there exists no such comparison study considering all these different approaches within the context of MATV for RTP. Therefore the objective of the present study was to provide a comprehensive comparison of the impact and accuracy of different methods taking into account the respiratory motion when defining the functional ITV in lung cancer RTP, by providing a framework consisting of state-of-the-art simulated Monte Carlo (MC) based datasets. The following strategies were compared: (a). delineating the GTV on a motion averaged image and assume it is the ITV [Caldwell 2003], (b). delineating the GTV on one respiratory frame and add fixed margins to derive the ITV, similarly to current practice using CT images [De Bari 2011], (c). considering the ITV as the union of the delineated GTVs on all respiratory frames, with or without incorporating motion correction based on different approaches.

## 2. Materials and methods

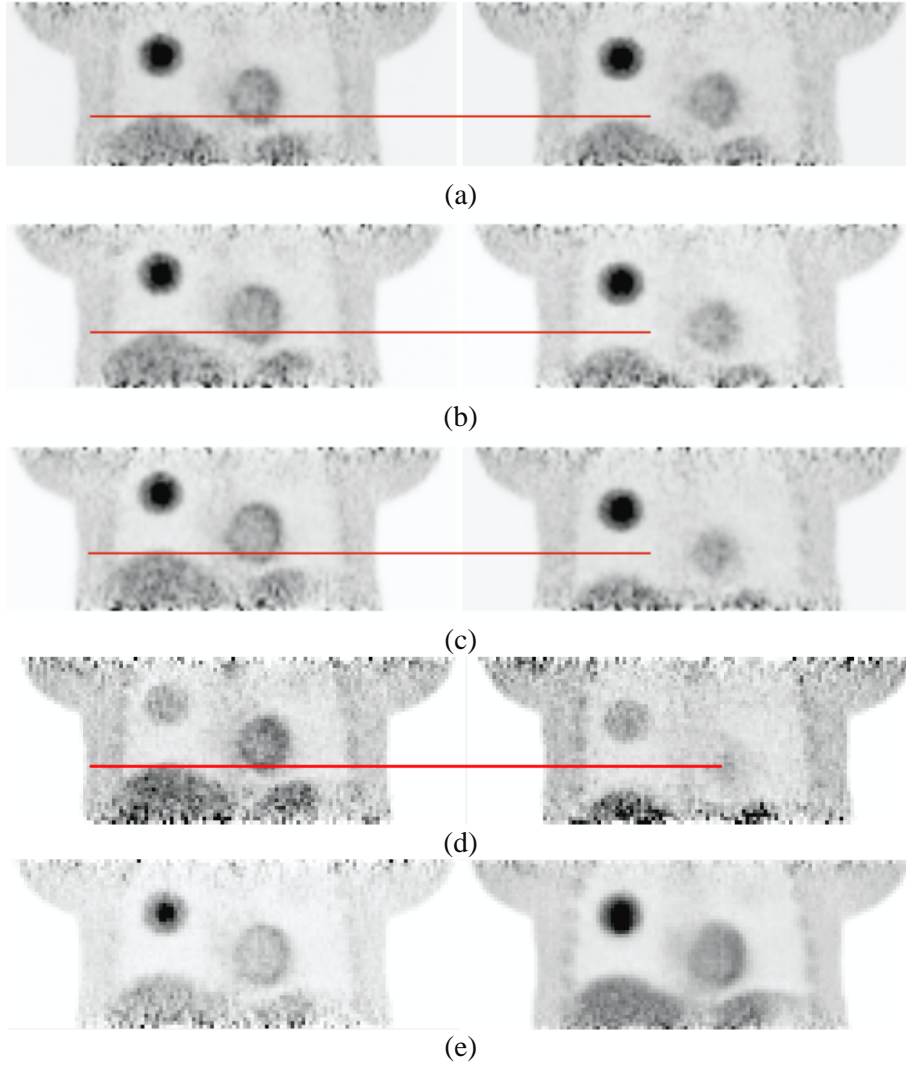
### 2.1 Evaluation datasets

Fifteen simulated datasets were generated using a model of the Philips GEMINI PET system implemented and validated [Lamare 2006] using the Monte-Carlo simulation platform in Geant4 Application for Tomography Emission (GATE) [Jan 2004], combined with the Non-Uniform Rational B-Spline (NURBS)-based 4D cardiac-torso phantom (NCAT) [Segars 2009]. Patient specific anatomical modifications were applied to the NCAT phantom as previously described [Le Maitre 2009], using patient-specific anatomy based on patient CT scans and associated respiratory signals. The simulation was carried out using the Real Time NCAT module implemented in GATE [Descourt 2006], with update of the phantom position every 0.1s. 150 positions of the phantom were simulated for each dataset, corresponding to three different respiratory cycles of 5s each (50 positions per 5s cycle). Finally, 3-minute acquisitions (36 respiratory cycles) were simulated in total. The list-mode data output of the simulation was subsequently divided in eight bins according to the amplitude of the respiratory motion.

The first 12 datasets correspond to the default anatomy of the NCAT with a regular sinusoidal respiratory motion. The phantom in figure 1 was simulated using three different maximum diaphragm displacements in the cranio-caudal direction (0.5, 1.5 and 2.5 cm, see figure 2), two different levels of tracer uptake heterogeneity within the tumor (either homogeneous or heterogeneous with a contrast between the intra-tumor activities set to 4), and two different tumor to background contrast (4:1 and 10:1). The tumor had a 5 cm diameter and the sub-volume a 2 cm diameter.



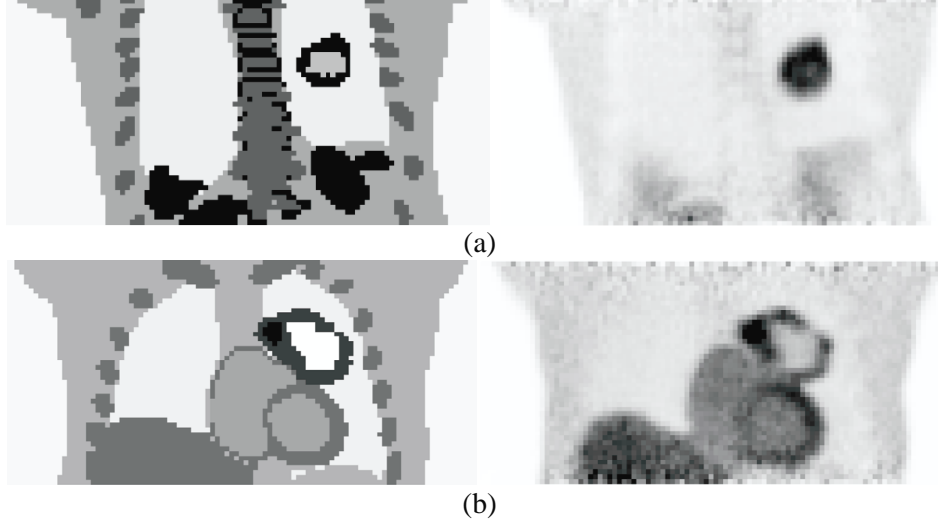
**Figure 1:** Phantom used to simulate the first 12 datasets. The inner tumor sphere (2cm diameter) has the same level of activity and four times the activity of the outer one for homogeneous (b). and heterogeneous (c). cases respectively.



**Figure 2:** Images simulated at the two extreme positions (left and right) of the respiratory cycle, using a maximum diaphragm displacement of (a) 0.5 cm, (b) 1.5 cm and (c-d) 2.5 cm. This phantom uses the NCAT default anatomy and a regular respiratory cycle. (a-c) contain a heterogeneous tumor, with a 10:1 tumor to background contrast ratio (and a contrast between the tumor and the inner sphere of 4:1), whereas (d) illustrates the 2.5 cm motion amplitude for a homogeneous case (contrast 4:1). The red line represents the reference diaphragm position. (e) shows the frame 1 (on the left) and the motion averaged image (on the right) for case (c).

For the 3 last datasets, both regular and irregular (in phase and in amplitude) motion were considered. The 13<sup>th</sup> dataset was simulated using the same phantom as the first 12 datasets with a heterogeneous tumor. The last two datasets were based on the NCAT phantom adapted to real PET/CT clinical datasets (figure 3). Two levels of motion amplitude were considered with maximum diaphragm displacement of 2.5 cm and 1.5 cm respectively. Different lesions based on manual segmentation of patients' <sup>18</sup>F-FDG images were incorporated in the each of these two datasets (14<sup>th</sup> and 15<sup>th</sup> dataset respectively). The first (see figure 3(a)) was an approximately spherical lesion (diameter of 4cm) with a tumor to background contrast of 20:1 and with a core (maximum length of 2 cm) exhibiting lower uptake with a contrast of 4:1 with respect to the lung. The second dataset (see figure 3(b)) had a heterogeneous, crescent-shaped tumor with a contrast of 2.5:1 and 9:1 between the high and low uptake regions and between the low uptake region and the background respectively. The

high uptake region was approximately spherical (1.8 cm diameter). This tumor had a necrotic core with activity similar as the background. For the 14<sup>th</sup> case (figure 3a), the core with lower uptake was considered part of the GTV ground-truth since its activity was above background, whereas in the 15<sup>th</sup> case (figure 3b) the necrotic core was considered outside the GTV ground-truth since its activity was the same as the background.



**Figure 3:** Illustration of the (a) 14<sup>th</sup> and (b) 15<sup>th</sup> datasets with patient-specific NCAT anatomy and realistic tumors: (left) the ground-truth and (right) the corresponding slice simulated PET image.

## 2.2 Respiratory motion incorporation and associated ITV definition

We compared three different strategies of taking into account respiratory motion within the ITV determination. The first strategy is similar to the clinical practice for morphological ITV definition on CT images without the use of a 4D-CT acquisition [de Bari 2011]. It consists in delineating the functional GTV on the mid-inspiration binned images and adding fixed margins (20mm and 10mm in the cranio-caudal and other directions respectively) in all binned images in order to account for the typical tumor displacement during the respiratory cycle allowing to derive the ITV. The other two approaches are in accordance with the recommendations of the AAPM task group number 76 [Keall 2006]. The second strategy consists in delineating the functional GTV on the motion averaged image, assuming it can be considered as the ITV since the motion averaged image is expected to contain the sum of all tumor positions during the breathing cycle [Caldwell 2003]. The last and third strategy requires delineating the MATV on all gated images, the final ITV being defined as the union of all delineated GTVs. For this last strategy, we considered four different approaches, two without and two with motion correction. Without motion correction, binned images were reconstructed with small ( $2 \times 2 \times 2 \text{ mm}^3$ ) and large voxels ( $4 \times 4 \times 4 \text{ mm}^3$ ). With motion correction, two different correction approaches (see section 2.3 below) were applied to the gated images prior to delineation in order to compare the results obtained using these correction methods, that also lead to different voxel size images. For these cases, list-mode data output was divided also in eight bins leading to eight synchronized PET images.

In all simulated datasets, volumes were automatically delineated using the Fuzzy Locally Adaptive Bayesian algorithm based on statistical and fuzzy modeling [Hatt 2009]. It is an accurate, robust and reproducible segmentation method, allowing the use of up to three classes for the delineation of heterogeneous tumors [Hatt 2010, 2011b]. Therefore, the entire volume delineated by the algorithm (either homogeneous volume or the union of the two classes in the heterogeneous tumor cases) was considered as the GTV.

This led to the comparison of the following six different reconstructed images:

- (A). *Frame 1 with margins*: GTV corresponding to the volume delineated on the mid-inspiration frame of the gated sequence reconstructed with standard  $4 \times 4 \times 4 \text{ mm}^3$  voxels.

To obtain the ITV fixed margins were added (20mm and 10mm in the cranio-caudal and other directions respectively),

- (B). *Motion averaged image*: GTV segmented on the motion averaged image reconstructed with standard  $4 \times 4 \times 4 \text{ mm}^3$  voxels and considered as the ITV,
- (C). a) *Union of LR images*: ITV corresponding to the union of GTVs delineated on each low resolution gated frame reconstructed with standard  $4 \times 4 \times 4 \text{ mm}^3$  voxels,  
 b) *Union of upsampled LR images*: ITV corresponding to the union of GTVs delineated on each low resolution gated frame reconstructed with smaller  $2 \times 2 \times 2 \text{ mm}^3$  voxels,  
 c) *Union of elastic-corrected images*: ITV corresponding to the union of GTVs delineated on each gated frame, reconstructed with the motion-incorporated reconstruction algorithm based on elastic transforms ( $4 \times 4 \times 4 \text{ mm}^3$  voxels),  
 d) *Union of SR-corrected images*: ITV corresponding to the union of GTVs delineated on each gated frame, reconstructed with the super-resolution reconstruction algorithm ( $2 \times 2 \times 2 \text{ mm}^3$  voxels).

### 2.3 Motion correction reconstruction algorithms

Two different reconstruction-incorporated motion correction methods (one based on elastic transformations [Lamare 2007b], the other on the super-resolution concept [Wallach 2010]) were considered here and compared regarding the third strategy in this study (union of GTVs delineated on gated frames). Although both motion correction approaches are applicable directly on reconstructed images [Lamare 2007b, Wallach 2012] or during the reconstruction process [Lamare 2007b, Wallach 2010], the second implementation approach using the simulated raw list-mode data, previously shown to lead to superior quality corrected images, was used in this work.

Let  $g_i, i = 1, \dots, I$  denote the counts detected in line of response (LOR)  $i$ ,  $y_j, j = 1, \dots, J$  the intensity of voxel  $j$  in the reconstructed image, and  $S_{ij}$  the probability of detecting at LOR  $i$  an event generated in voxel  $j$ . The static tomographic problem is:

$$g = Sy. \quad (1)$$

In order to adapt equation (1) to the case of reconstruction with respiratory motion, we introduce the index  $k = 1, \dots, K$ , representing the respiratory phase,  $g_{ik}$  the counts detected in LOR  $i$  during respiratory state  $k$ ,  $y_{jk}$  the intensity of voxel  $j$  during respiratory state  $k$ , and  $S_{kij}$  the probability of detecting at LOR  $i$  an event generated in voxel  $j$  during respiratory state  $k$ . Equation (1) hence becomes:

$$\begin{aligned} g_k &= S_k y_k \\ &= S_k M_k y, \end{aligned} \quad (2)$$

where  $y$  is the reconstructed image in the reference position, and  $M_k$  is the motion matrix between the reference position and position  $k$ . The model (2) was used to build both motion-incorporated reconstruction algorithms used in our study.

The first correction method is based on the incorporation of elastic transformations into the one-pass list mode EM (OPL-EM) reconstruction algorithm [Lamare 2007b]. Its update equation is:

$$y_j^{n+1} = \frac{y_j^n}{\sum_k \sum_i (S_k M_k)_{ij}} \sum_{k=1}^K \sum_{i \in T^n} (S_k M_k)_{ij} \frac{1}{(S_k M_k y^n)_i}, \quad (3)$$

where  $n$  is both the iteration number and the subset used in that iteration.  $T^n$  is the set of list-mode events in the  $n^{\text{th}}$  subset. Images reconstructed using this correction scheme have the same voxel size as the original images.

The second reconstruction scheme incorporates a super-resolution (SR) model. SR techniques make use of the motion in a sequence of low-resolution (LR) images in order to improve them [Kang 2003]. By modeling each image of the original sequence as a degraded version of an underlying high-resolution (HR) image with wider bandwidth than that of any of the individual LR frames, it is possible to estimate the high-resolution image. Let us consider the 3D HR image of size  $LJ$ , written as the vector  $f$ , which is the ideal image resulting from sampling (at or above the Nyquist rate) a continuous, band-limited 3D scene. The parameter  $L$  represents the down-sampling factor in the observation model. Each of the  $K$  observed LR images is of size  $J$  and is denoted in lexicographical notation as  $y_k, k = 1$  to  $K$ . Using these notations, a typically used degradation model is:

$$y_k = DBM_k f + \varepsilon_k, \quad k = 1..K, \quad (4)$$

where  $D$  is a  $J \times LJ$  subsampling matrix performing an averaging over neighboring voxels,  $B$  is a  $LJ \times LJ$  blur matrix that models the action of the point spread function of the PET system,  $M_k$  is the same  $LJ \times LJ$  warp matrix as in equation (2), and  $\varepsilon_k$  is an additive noise vector. This degradation model can be incorporated into the tomographic problem [Wallach 2010] incorporating motion in equation (2), yielding:

$$\begin{aligned} g_k &= S_k y_k \\ &= S_k DBM_k f + \varepsilon_k. \end{aligned} \quad (5)$$

This model was incorporated within the one-step late (OSL) algorithm [Green 1990], a maximum a posteriori (MAP)-based reconstruction method:

$$f_j^{n+1} = \frac{f_j^n}{\sum_k \sum_i (S_k DBM_k)_{ij} + \lambda \frac{\partial V}{\partial f_j}(f^n)} \sum_{k=1}^K \sum_{i=1}^I (S_k M_k)_{ij} \frac{g_{ik}}{(S_k DBM_k f^n)_i}, \quad (6)$$

where,  $V$  is a regularization term representing *a priori* information about the image  $f$ . We chose the following form for function  $V$

$$V(f) = \sum_{j=1}^J \sum_{l \in \mathcal{N}_j} V_\delta(f_j - f_l), \quad (7)$$

where,  $\mathcal{N}_j$  is the neighborhood of voxel  $j$  and  $V_\delta$  is the Huber function [Huber 1981]

$$V_\delta(z) = \begin{cases} z^2 & \text{if } |z| < \delta \\ 2\delta|z| - \delta^2 & \text{if } |z| \geq \delta. \end{cases} \quad (8)$$

$\delta$  is a positive parameter controlling the switch between the quadratic and the linear part of the function. This function is quadratic near the origin and linear far from the origin, which means that small intensity differences between neighboring voxels are more penalized than large ones, assumed to be associated with signal rather than noise. It should be noted that the images reconstructed using this second method have voxels twice as small as the original images.

In both reconstruction methods (3) and (6), the system matrices  $S_k$  include normalization and attenuation corrections and are computed as follows:

$$S_k = N A_k P_k. \quad (9)$$

$N$  is a diagonal  $I \times I$  matrix whose term  $N_{ii}$  is the normalization term of the LOR  $i$  and  $P_k$  is a  $I \times J$  matrix whose term  $P_{kij}$  represents the probability of detecting at LOR  $i$  an event generated in voxel  $j$  during respiratory state  $k$ , considering only geometric effects such as the length of intersection between LOR  $i$  and voxel  $j$ . The diagonal matrix  $A_k$  is of size  $I \times I$  and models the attenuation occurring along the LOR  $i$

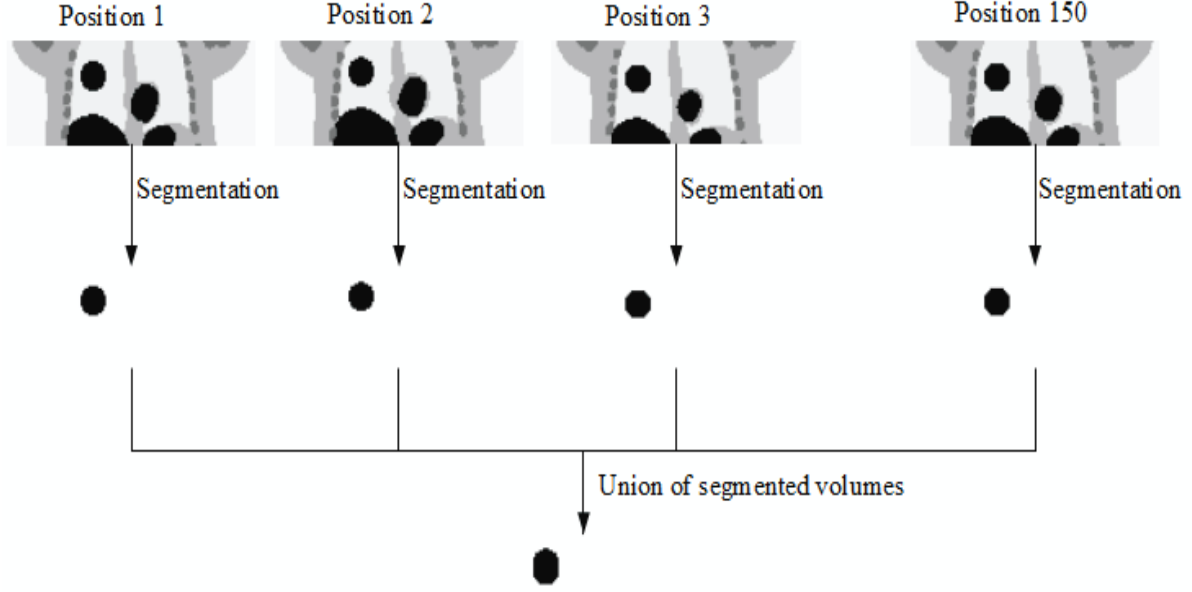
$$A_{kii} = \exp\left(-\sum_{j=1}^J P_{kij}(M_k \mu)_j\right), \quad (10)$$

where  $\mu_j$  is the linear attenuation coefficient at the energy of 511 keV ( $\mu_j$  is the intensity of the voxel  $j$  in the NCAT attenuation image).

#### 2.4 Data analysis and figures of merit

The six results (A), (B), (C)-a), (C)-b), (C)-c), (C)-d), described in section 2.2 above were compared in terms of accuracy with respect to the ground-truth ITV, known from the NCAT simulated phantoms. This ground-truth ITV was defined as the union of all ground-truth GTVs in the 150 ground-truth images generated with the NCAT phantom (figure 4).





**Figure 4:** Definition of the ground-truth ITV as the union of all ground-truth GTVs.

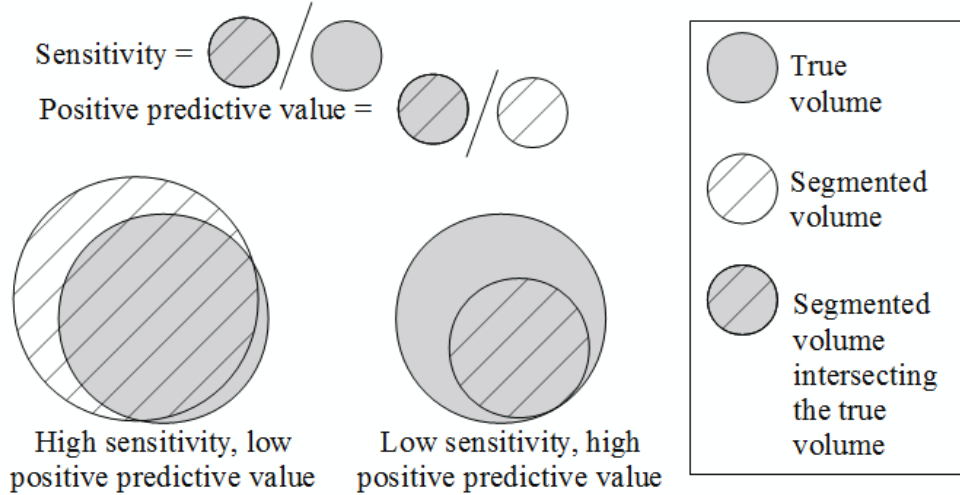
The performance of segmentation was assessed using the combination of two figures of merit, namely the sensitivity (SE) and the positive predictive value (PPV). Let us denote  $A_T$  the true lesion area,  $A_M$  the segmented area,  $A_T \cap A_M$  the intersection between areas  $A_T$  and  $A_M$ , and  $\text{size}(A)$  the size of area  $A$ . Sensitivity (SE) is defined as the ratio between the size of the segmented area intersecting the true area and the true size:

$$\text{sensitivity} = \frac{\text{size}(A_T \cap A_M)}{\text{size}(A_T)}. \quad (12)$$

Positive predictive value (PPV) is defined as the ratio between the size of the segmented area intersecting the true area and the measured size:

$$\text{positive predictive value} = \frac{\text{size}(A_T \cap A_M)}{\text{size}(A_M)}. \quad (13)$$

Using both the SE and PPV, the delineation accuracy can be assessed (figure 5). The combination of a high PPV and a low SE indicates an under-evaluation of the true volume, whereas the inverse indicates an over-evaluation of the true volume. Since the combination of high SE and PPV indicates an accurate delineation the average of SE and PPV were also calculated.



**Figure 5:** Definitions of sensitivity (SE) and positive predictive value (PPV).

Statistical tests were carried out with Medcalc<sup>TM</sup> (Medcalc software, Belgium) to assess differences between SE, PPV, and average of SE and PPV calculated for all six results, using Mann-Whitney U or Kruskal-Wallis tests for two, three or more distributions respectively. All tests were two-sided and a p value below 0.05 was considered statistically significant.

### 3. Results

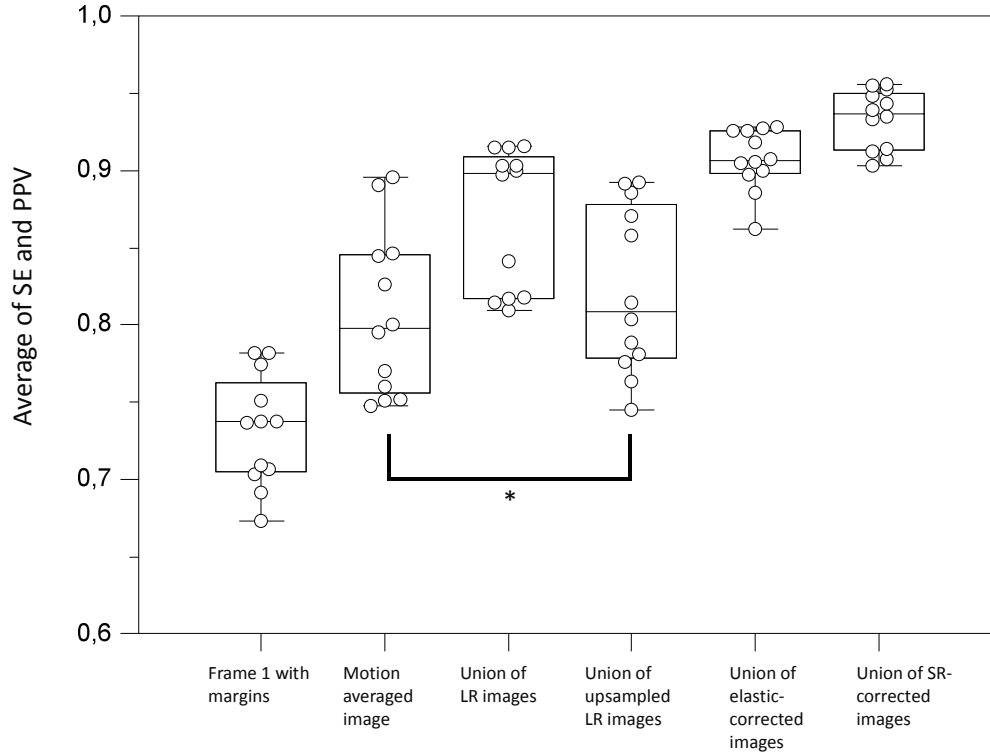
Results are presented separately for the first twelve standard NCAT datasets with a regular respiratory motion only (section 3.1), and for the three other datasets (realistic NCAT modified phantoms), considering both regular and irregular motion (section 3.2).

#### 3.1 Standard NCAT cases

Values of SE and PPV are presented in table 1 whereas figure 6 shows the values of averaging PPV and SE.

n=12	Mean±SD
<i>Sensitivity</i>	
Frame 1 with margins	0.989±0.016
Motion averaged image	0.704±0.109
Union of LR images	0.907±0.033
Union of upsampled LR images	0.824±0.104
Union of elastic-corrected images	0.912±0.016
Union of SR-corrected images	0.931±0.015
<i>Positive predictive value</i>	
Frame 1 with margins	0.475±0.087
Motion averaged image	0.920±0.052
Union of LR images	0.830±0.087
Union of upsampled LR images	0.823±0.041
Union of elastic-corrected images	0.903±0.042
Union of SR-corrected images	0.934±0.042

**Table 1:** SE and PPV for the compared ITVs of all 12 datasets with a regular breathing motion.



**Figure 6:** Distributions of the average of SE and PPV values for the compared ITVs of all 12 datasets with a regular breathing motion. \* denotes non-significant difference between distributions.

According to the average of SE and PPV values (figure 6), the results point to an advantage of using the two motion-corrected images, resulting in values between 0.9 and 1, with a slight but significant advantage for SR-corrected images over elastic-corrected images ( $p=0.0056$ ). On the one hand, no significant differences were seen between the two correction methods regarding PPV, although SR-corrected images led to slightly higher values ( $0.90\pm0.04$  vs.  $0.93\pm0.04$ ,  $p=0.052$ ). On the other hand, SE was significantly higher for SR-corrected images ( $0.91\pm0.02$  vs.  $0.93\pm0.02$ ,  $p=0.01$ ) (table 1).

All methods led to significantly different results of average SE and PPV, except between upsampled LR images and motion averaged image ( $p=0.45$ ). The worst performance was obtained by delineation on the first gated PET frame with the addition of fixed margins, with an average of SE and PPV between 0.65 and 0.8. This approach led to significantly higher SE (close to 1) but significantly lower PPV ( $<0.5$ ) than all other 5 results (table 1) ( $p<0.0001$ ). The delineation performed on the motion averaged image led to the second most inaccurate volume, with mean SE of 0.7 (significantly lower than all other approaches,  $p<0.0001$ ) and PPV of 0.9 (significantly higher than all approaches except the images corrected for motion,  $p<0.01$ ) respectively. All other approaches dealing with LR images (for both voxel sizes) or corrected images (using either elastic transforms or SR) resulted for all cases in both mean SE and PPV above 0.8. The use of smaller voxels in the reconstruction of LR images led to significantly lower SE ( $0.91\pm0.03$  vs.  $0.82\pm0.10$ ,  $p=0.04$ ) but similar PPV ( $0.83\pm0.09$  vs.  $0.82\pm0.04$ ,  $p=0.4$ ).

#### Effects of motion amplitude

Figures of merit are presented for each motion amplitude (0.5, 1.5 and 2.5 cm) in table 2 and figure 7(a-c).  $P$  values are provided in table 2 (values  $<0.05$  in **bold**) to demonstrate statistical significance of the sensitivity to increasing motion amplitude.

	n=4	Mean $\pm$ SD				
Sensitivity	Amplitude	Amplitude	$p$ r/t*	Amplitude	$p$ r/t*	$p$ r/t*
	0.5 cm	1.5 cm	0.5 cm	2.5 cm	0.5 cm	1.5 cm

Frame 1 with margins	1.00±0.001	0.996±0.001	<b>0.02</b>	0.974±0.018	<b>0.02</b>	<b>0.02</b>
Motion averaged image	0.833±0.017	0.669±0.032	<b>0.02</b>	0.602±0.043	<b>0.02</b>	0.08
Union of LR images	0.922±0.023	0.915±0.011	0.4	0.885±0.045	0.3	0.4
Union of upsampled LR images	0.844±0.109	0.837±0.069	0.8	0.794±0.138	0.4	0.8
Union of elastic-corrected images	0.916±0.011	0.901±0.006	<b>0.04</b>	0.916±0.024	0.6	0.2
Union of SR-corrected images	0.942±0.011	0.936±0.007	0.2	0.915±0.009	<b>0.03</b>	<b>0.04</b>
<i>Positive predictive value</i>						
Frame 1 with margins	0.390±0.031	0.468±0.045	<b>0.02</b>	0.565±0.050	<b>0.02</b>	<b>0.04</b>
Motion averaged image	0.907±0.062	0.914±0.067	0.8	0.937±0.041	0.4	0.2
Union of LR images	0.802±0.108	0.841±0.099	0.8	0.850±0.073	0.8	0.6
Union of upsampled LR images	0.811±0.039	0.811±0.058	0.8	0.844±0.031	0.3	0.3
Union of elastic-corrected images	0.911±0.026	0.920±0.043	0.4	0.883±0.055	0.6	0.3
Union of SR-corrected images	0.923±0.047	0.939±0.048	0.4	0.942±0.044	0.8	0.6

**Table 2:** SE and PPV for the compared ITVs of all 12 standard NCAT datasets with a regular breathing motion with respect to motion amplitude. \*  $p$  values are provided for significance of the dependency of SE and PPV for the 1.5cm amplitude relative to (r/t) the 0.5 cm one, and for 2.5cm relative to 0.5 and 1.5cm.

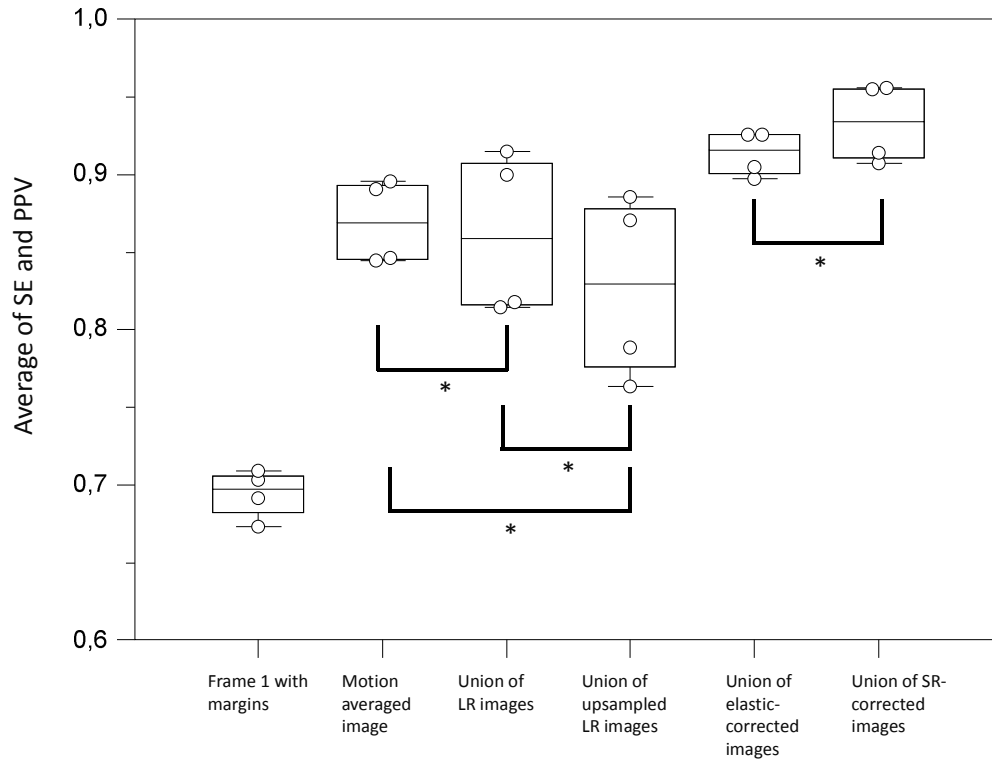
Regarding SE, the two strategies showing the highest dependency to motion amplitude were Frame 1 with margins and motion averaged image, although the union of SR-corrected images also exhibited significant decrease of SE with increasing motion amplitude. It should be emphasized however that whereas the union of SR-corrected images and Frame 1 with margins resulted in very high SE ( $>0.9$ ) for all motion amplitudes, in the case of the motion averaged image, the SE decreased from satisfactory values ( $0.83\pm0.02$ ) to significantly lower SE values of  $0.67\pm0.03$  and  $0.60\pm0.04$  for motion amplitudes of 1.5 and 2.5 cm respectively ( $p=0.02$ ).

Regarding PPV, the only approach that exhibited significantly increased values (from  $0.39\pm0.03$  to  $0.47\pm0.05$  and  $0.57\pm0.05$  for 1.5 and 2.5 motion amplitude respectively), was Frame 1 with margins, although the absolute PPV values were low ( $<0.6$ ) even for high amplitude.

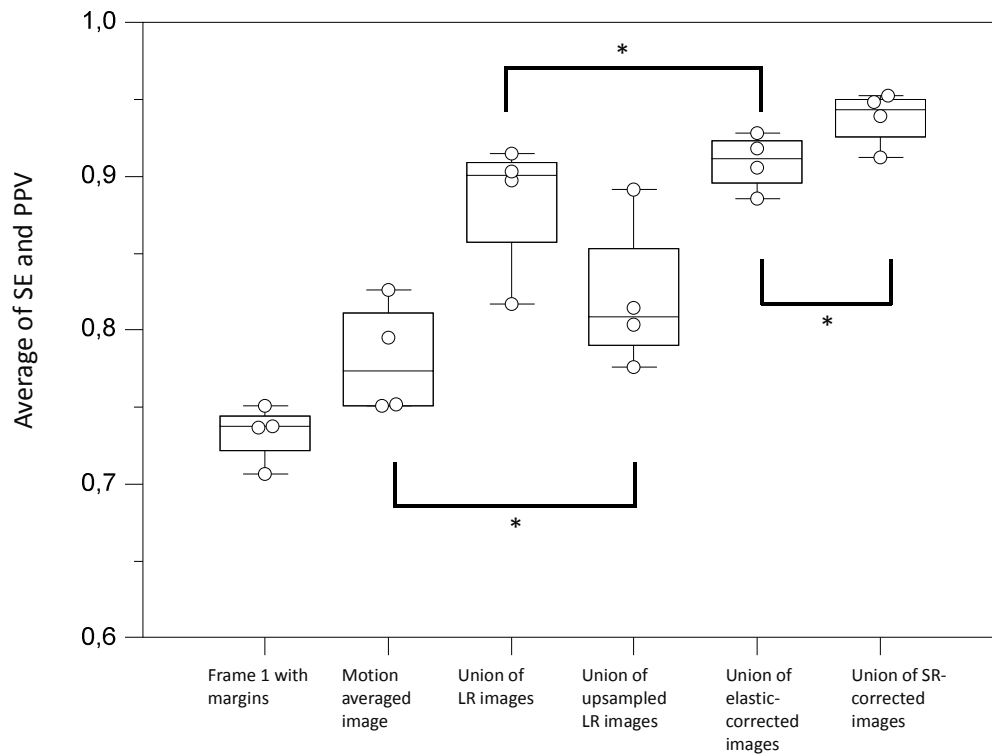
The performance hierarchy of the different motion approaches considered was affected by increasing motion amplitudes (figure 7). For the minimum motion amplitude (0.5 cm, figure 7a) there was no significant difference between the performance of the use of the motion averaged image, LR images and upsampled LR images, although there was a trend of superior performance considering the motion averaged image. With increasing motion amplitude, the motion averaged image performance tended to worsen, whereas performance of LR images improved with significantly higher values than motion averaged image and upsampled LR images (figures 7b and 7c). Frame 1 with margins led to significantly worse performance than all other approaches, except for a motion amplitude of 2.5cm for which a similar performance to the motion averaged image was obtained (figure 7c), considering that the performance of Frame 1 with margins improved with increasing motion amplitudes (see figures 7a-c).

The two results based on motion corrected images led to systematically significantly better performance than the other methods for all motion amplitudes (figures 7a-c). Only a small advantage, although not statistically significant ( $p=0.08$ ) was observed for the SR-corrected over the elastic-corrected images, especially for 1.5 and 2.5 cm motion amplitudes. SR-corrected images led to significantly higher performance than all other methods for all motion amplitudes. On the contrary, elastic-corrected images led to significantly higher performance than all methods except with respect to LR images for motion amplitudes higher than 0.5 cm, for which only a non-significant performance improvement was observed.

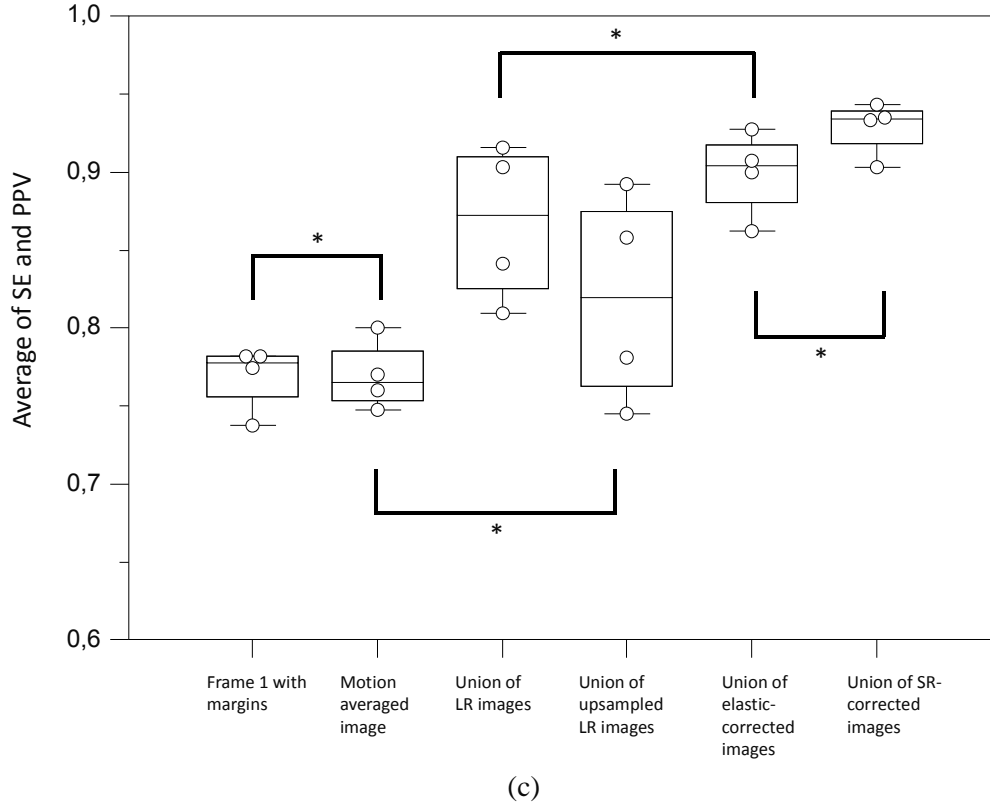
# Incorporation of motion information for tumour volume definition in PET



(a)



(b)



**Figure 7:** Distributions of the average of SE and PPV for the compared ITVs of all datasets with a regular breathing motion with respect to different motion amplitudes, (a) 0.5, (b) 1.5 and (c) 2.5cm. \* denotes non-significant difference between distributions.

#### *Effect of tumor to background contrast*

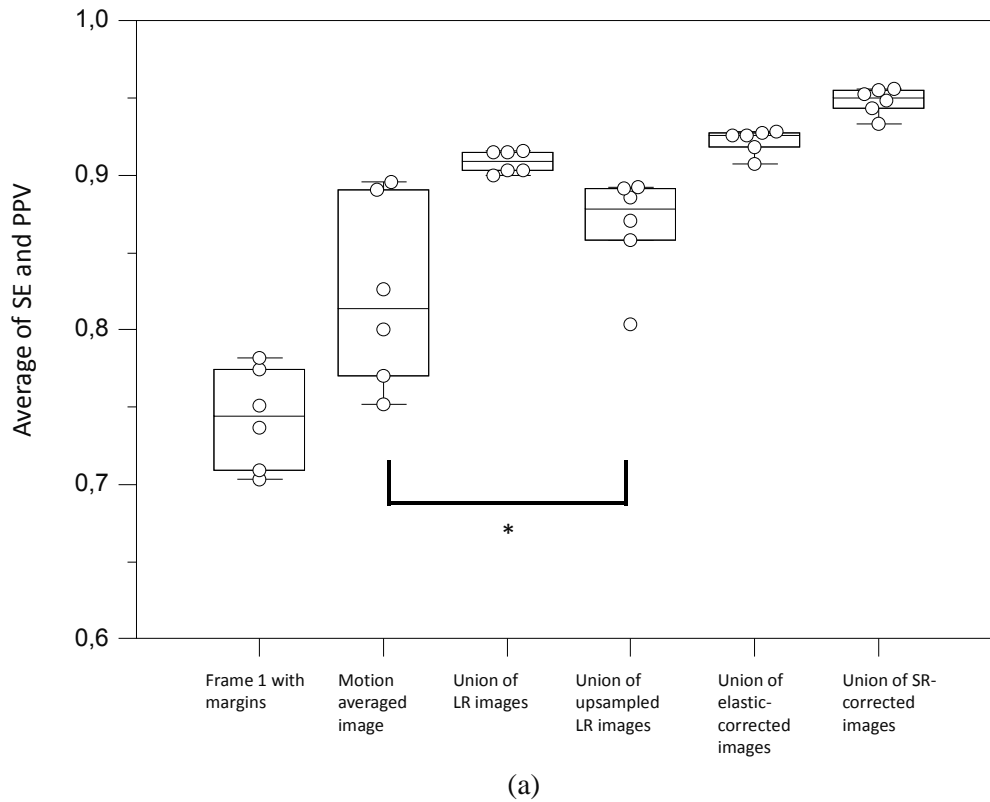
Figures of merit are presented for each tumor to background contrast considered (4:1 and 10:1) in order to illustrate the dependency on the contrast in table 3 and figures 8(a-c).

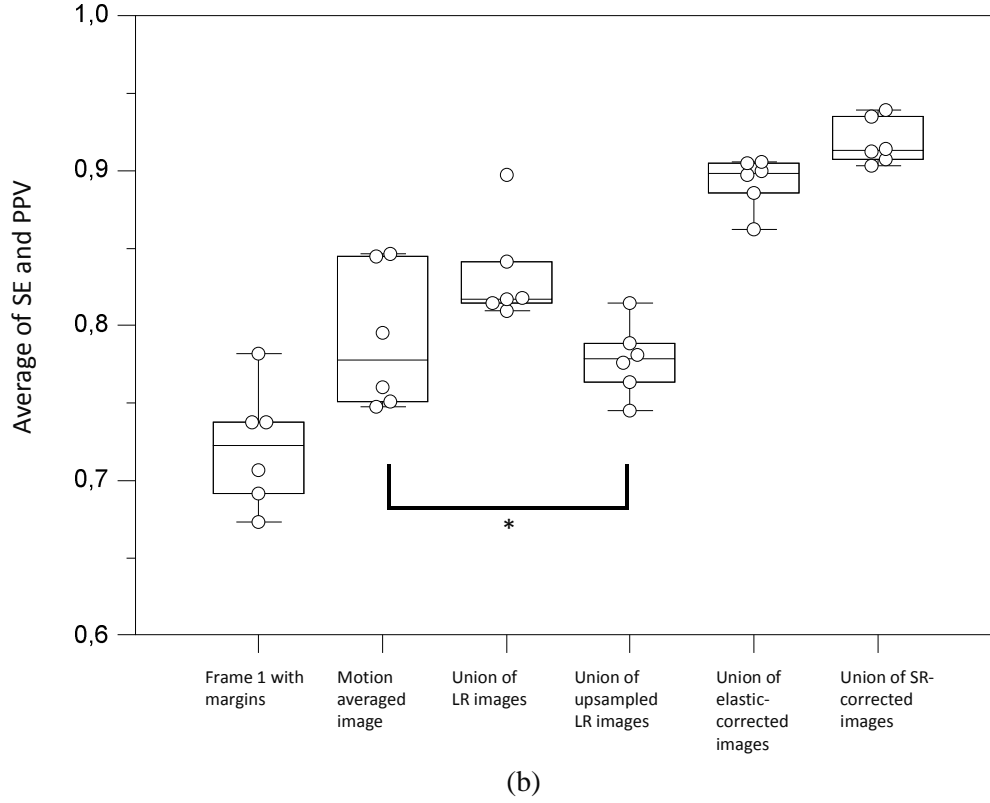
n=6	Mean±SD		
<i>Sensitivity</i>	Contrast 10:1	Contrast 4:1	<i>p</i> r/t 10:1*
Frame 1 with margins	0.994±0.008	0.985±0.020	0.4
Motion averaged image	0.694±0.112	0.733±0.112	0.9
Union of LR images	0.921±0.008	0.890±0.040	0.09
Union of upsampled LR images	0.902±0.033	0.747±0.078	<b>0.007</b>
Union of elastic-corrected images	0.909±0.009	0.910±0.024	0.6
Union of SR-corrected images	0.930±0.018	0.928±0.015	0.8
<i>Positive predictive value</i>			
Frame 1 with margins	0.492±0.072	0.413±0.137	0.5
Motion averaged image	0.952±0.047	0.870±0.036	<b>0.04</b>
Union of LR images	0.896±0.019	0.736±0.069	<b>0.01</b>
Union of upsampled LR images	0.832±0.042	0.797±0.052	0.3
Union of elastic-corrected images	0.935±0.017	0.849±0.046	<b>0.01</b>
Union of SR-corrected images	0.966±0.010	0.890±0.036	<b>0.01</b>

**Table 3:** SE and PPV for the compared ITVs of all datasets with a regular breathing motion with respect to tumor to background contrast. \* *p* values are provided for significance of the dependency of SE and PPV relative to (r/t) contrast value.

Regarding SE, only upsampled LR images exhibited significant dependency to decreasing tumor contrast ( $0.90 \pm 0.03$  vs.  $0.75 \pm 0.08$ ,  $p=0.007$ ). On the contrary, regarding PPV, all methods, except upsampled LR images and Frame 1 with margins, exhibited significantly lower values for lower contrast ( $p$  0.04-0.01) although they also showed decreasing values.

The hierarchy between the methods was not significantly modified by the decreasing contrast, with only the motion averaged image and upsampled LR images resulting in non-significantly different performances for both contrast configurations (figure 8). The performance of LR images was significantly reduced with lower contrast (from  $>0.9$  to  $0.8-0.9$ ) but stayed significantly higher than the motion averaged image and upsampled LR images. Frame 1 with fixed margins was the worst performing method, whereas the top two methods were consistently the motion-corrected images, with a significant advantage of the union of SR-corrected images, in both contrast configurations.





**Figure 8:** Distributions of the average of SE and PPV for the compared ITVs of all datasets with a regular breathing motion with respect to (a) 10:1 and (b) 4:1 contrast. \* denotes non-significant difference between distributions.

### 3.2 Realistic NCAT cases

This dataset contains three different phantoms, one based on the standard NCAT anatomy with a “simple” heterogeneous lesion (one sphere within another), whereas the two others are based on NCAT phantoms adapted to specific anatomies of real patients, and more realistic lesions in terms of both shape and heterogeneity uptake distributions (section 2.1, figure 3). For these three phantoms, both regular and irregular respiratory motion was simulated, with two different amplitudes of maximum 1.5 and 2.5 cm respectively.

#### Regular vs. irregular motion

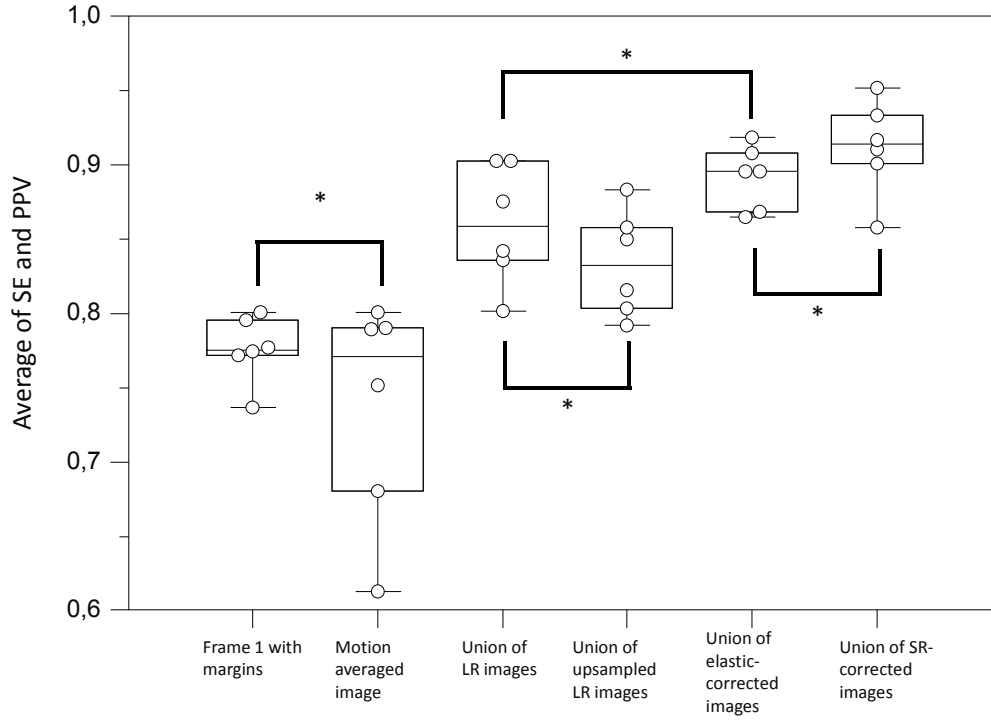
Values of SE and PPV for these three simulated datasets, for both regular and irregular motion, are presented in table 4 whereas figure 9(a-b) shows the values of averaging PPV and SE.

n=6	Mean±SD		
	Regular motion	Irregular motion	<i>p</i> r/t regular*
Sensitivity			
Frame 1 with margins	0.935±0.045	0.940±0.051	0.9
Motion averaged image	0.564±0.140	0.509±0.148	0.2
Union of LR images	0.837±0.071	0.762±0.030	<b>0.016</b>
Union of upsampled LR images	0.826±0.047	0.755±0.046	<b>0.025</b>
Union of elastic-corrected images	0.848±0.043	0.787±0.015	<b>0.0039</b>
Union of SR-corrected images	0.860±0.066	0.799±0.052	0.07
Positive predictive value			
Frame 1 with margins	0.618±0.085	0.600±0.048	0.4

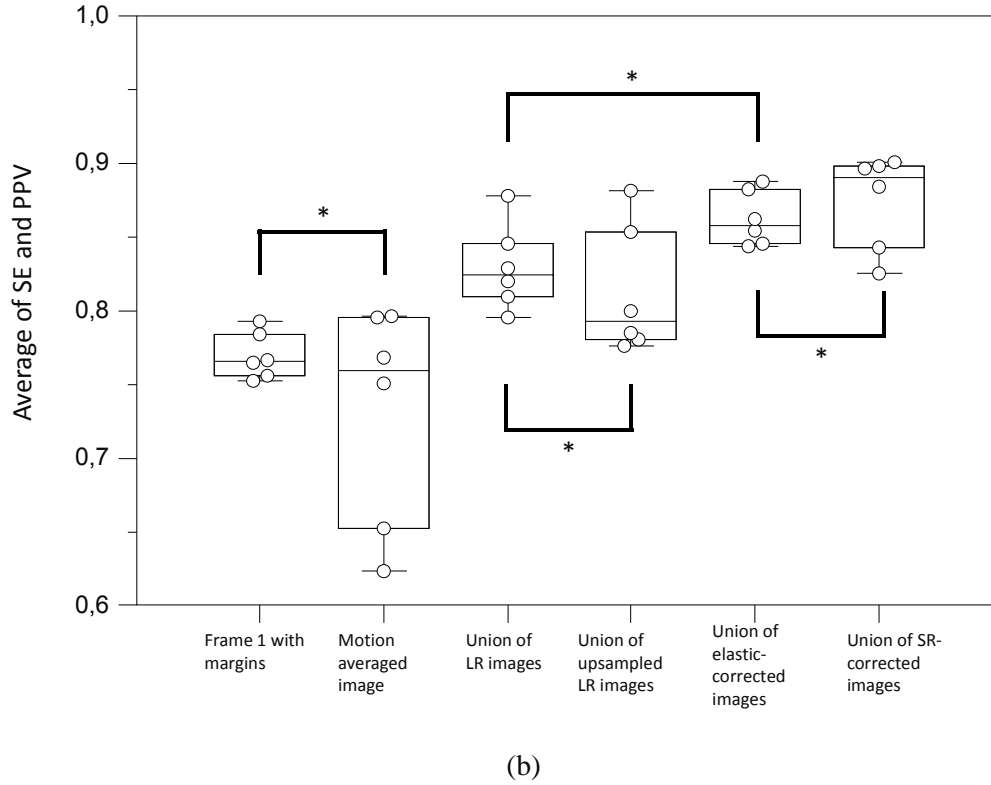


Motion averaged image	0.912±0.039	0.954±0.044	0.15
Union of LR images	0.883±0.052	0.897±0.062	0.4
Union of upsampled LR images	0.842±0.060	0.871±0.051	0.3
Union of elastic-corrected images	0.935±0.027	0.938±0.028	0.9
Union of SR-corrected images	0.964±0.010	0.950±0.033	0.3

**Table 4:** SE and PPV for the compared ITVs of all datasets 13<sup>th</sup> to 15<sup>th</sup> for a regular vs. irregular breathing motion. \*  $p$  values are provided for significance of the dependency of SE and PPV relative to (r/t) regularity of breathing.



(a)



**Figure 9:** Distributions of the average of SE and PPV for the compared ITVs of all datasets with respect to (a) regular, and (b) irregular breathing motion. \* denotes non-significant difference between distributions.

Regarding SE, all methods led to decreased sensitivity with irregular motion, although it was significant only for the union of LR (upsampled or not) and the elastic-corrected images ( $p < 0.03$ ). SR-corrected images exhibited also lower sensitivity with irregular motion but the trend was not significant ( $p = 0.07$ ). Frame 1 with margins and the motion averaged image led to very low SE values ( $< 0.7$ ), irrespective of regular or irregular motion ( $p > 0.2$ ). On the contrary, although PPV values were systematically higher, none of the methods exhibited significantly different values for irregular breathing motion ( $p > 0.1$ ).

The hierarchy between the methods was similar to that observed on the first dataset of standard NCAT phantoms, and was not impacted by the regular nature of breathing motion (figure 9). The best performance for regular ( $0.912 \pm 0.032$ ), as for irregular ( $0.875 \pm 0.032$ ) motion was observed for the ITVs delineated on SR-corrected images. Although there was a trend of overall slightly higher performance for the use of SR-corrected over elastic-corrected images, especially for regular motion ( $0.892 \pm 0.021$  vs.  $0.912 \pm 0.032$ ,  $p = 0.2$ ), this trend was not statistically significant ( $p > 0.2$ ). In addition, there was also no statistically significant difference between the performance of Frame 1 with margins and motion average image, although the latter exhibited a larger spread of values (0.6-0.75) whereas Frame 1 with margins had a much smaller spread with all values between 0.75 and 0.8. Finally, there was no statistically significant difference between the standard LR and upsampled LR images, with a performance situated between, on one hand, Frame 1 with margins and motion average image, and on the other hand, the motion corrected images (0.75-0.9).

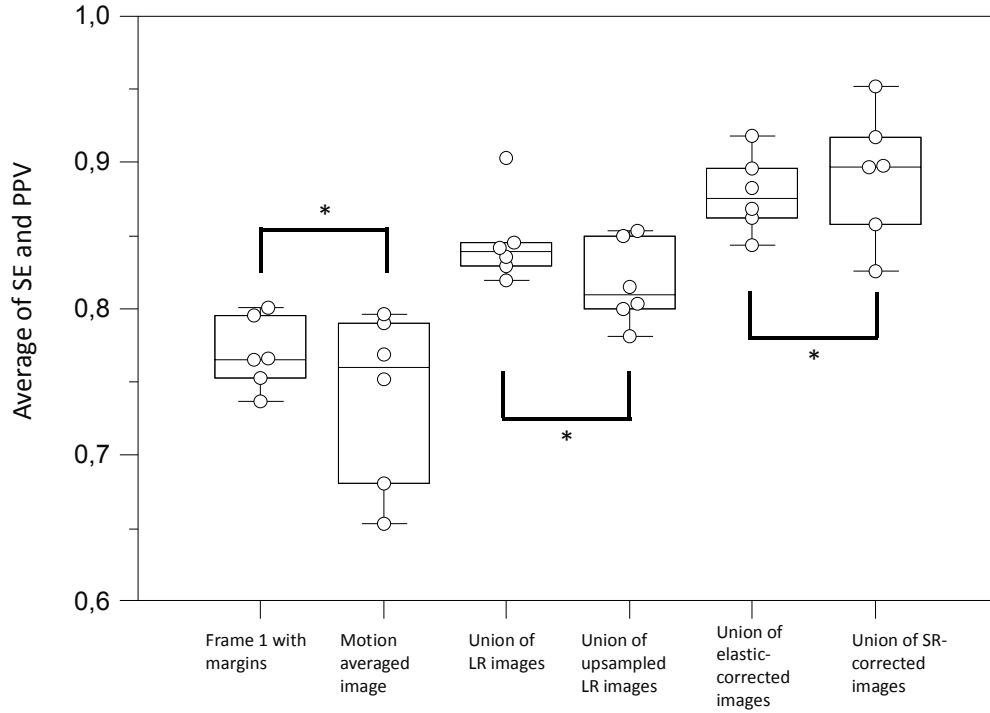
#### *Varying motion amplitude*

Values of SE and PPV for the three phantoms, for different motion amplitudes, are presented in table 5 whereas figure 10(a-b) shows the values of averaging PPV and SE.

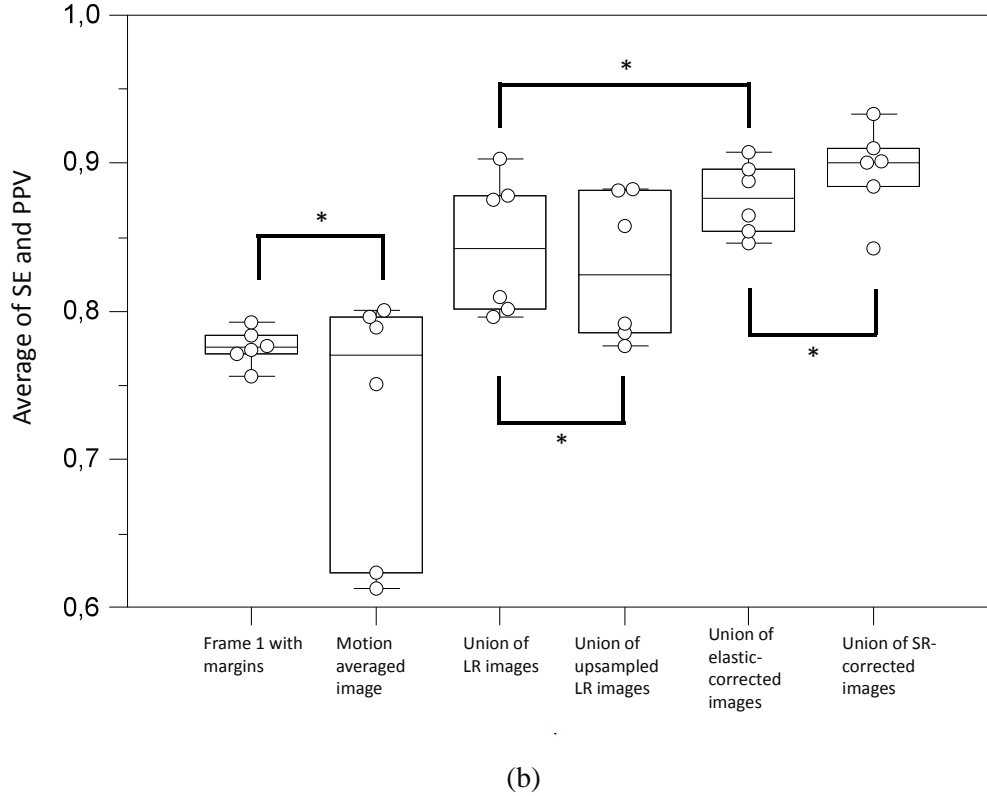
n=6	Mean±SD
-----	---------

	Amplitude 1.5 cm	Amplitude 2.5 cm	$p$ r/t 1.5cm*
<b>Sensitivity</b>			
Frame 1 with margins	0.932±0.052	0.943±0.044	0.9
Motion averaged image	0.560±0.093	0.513±0.184	1.0
Union of LR images	0.794±0.069	0.805±0.067	0.3
Union of upsampled LR images	0.790±0.040	0.791±0.076	0.9
Union of elastic-corrected images	0.815±0.044	0.820±0.049	0.9
Union of SR-corrected images	0.830±0.074	0.830±0.061	0.9
<b>Positive predictive value</b>			
Frame 1 with margins	0.609±0.030	0.608±0.094	0.6
Motion averaged image	0.921±0.046	0.945±0.045	0.3
Union of LR images	0.898±0.020	0.883±0.079	0.6
Union of upsampled LR images	0.845±0.051	0.868±0.061	0.8
Union of elastic-corrected images	0.942±0.026	0.932±0.028	0.6
Union of SR-corrected images	0.953±0.028	0.961±0.021	0.7

**Table 5:** SE and PPV for the compared ITVs of datasets 13<sup>th</sup> to 15<sup>th</sup> for a breathing motion amplitude of 1.5 and 2.5 cm. \*  $p$  values are provided for significance of the dependency of SE and PPV relative to amplitude of motion.



(a)



**Figure 10:** Distributions of the average of SE and PPV for the compared ITVs of all datasets with respect to (a) 1.5 cm, and (b) 2.5 cm breathing motion amplitude. \* denotes non-significant difference between distributions.

There was no statistically significant dependency on the amplitude of motion for neither SE nor PPV and for none of the six ITVs under comparison. The hierarchy between the six different approaches considered was not significantly impacted by the motion amplitude, except for the difference between elastic-corrected and LR images, which was significant only for the 1.5cm motion amplitude.

#### 4. Discussion

Respiratory motion still represents a major limitation of the accuracy with which functional parts of the target volume can be delineated on PET images for treatment planning in current clinical practice. Several studies have addressed this issue within the context of ITV determination in order to take into account the motion [Riegel 2012, Caldwell 2003, Nehmeh 2004, Wolthaus 2005, Aristophanous 2011, Aristophanous 2012], suggesting for instance that tumor volumes defined as the union of 4D gated PET frames may better define the physiological extent of moving tumors and improve radiation treatment planning for lung tumors [Aristophanous 2012]. Most of these studies were carried out on clinical patient images without a known ground-truth ITV and without applying motion correction which can be used to improve the quality of gated frame images. Another recent study proposed to derive the ITV from fixed threshold-based segmentations of MIP projections in 4D PET gated frames [Lamb 2011]. The proposed approach was carried out on only 3 patients using 4D-CT as the reference for the known ITV.

The purpose of the current study was to provide a framework and results for a full comparison of three different strategies allowing to incorporate temporal information into the delineation of the internal target volume from PET images. These strategies included adding margins to the volume defined on a single binned image, segmenting a motion averaged image assumed to contain all positions of the tumor throughout the respiratory cycle, and thirdly using the union of GTVs delineated on binned frames. The issue associated with the use of this latter strategy is that respiratory binning in PET yields images of reduced contrast and signal to noise ratio, as each image only

contains a fraction of the overall acquired data. We therefore considered the use of two different motion correction methods, both incorporated into the image reconstruction step. The first one is based on the incorporation of elastic transformations into the OPL-EM reconstruction algorithm [Lamare 2007b]. The second one incorporates a super-resolution model in the OSL reconstruction algorithm [Wallach 2010]. Both enable the reconstruction of one frame of the binned sequence with the statistics corresponding to the entire acquisition. The ITVs obtained as the union of the volumes delineated on each binned image thus reconstructed were compared to those obtained as the union of the volumes delineated on non-corrected binned frames, reconstructed with two different voxel sizes. Contrary to most of the studies previously conducted in the subject, all volumes were automatically segmented using a previously validated automatic segmentation based on statistical and fuzzy modeling [Hatt 2009, 2010, 2011a] instead of fixed thresholding.

In this study we decided not to work with clinical images for which it is difficult to obtain the ground-truth. We instead relied on very realistic simulated data generated using the GATE simulation platform in combination with the NCAT phantom and a previously validated model of the Philips GEMINI scanner. We considered both regular and irregular respiratory motion, with realistic motion amplitudes ranging from 0.5 cm to 2.5 cm.

Adding margins to the volume defined on a single binned image led to an ITV much larger than the true one, covering it almost completely, as measured by a very high SE ( $>0.9$ ), but also associated with a low PPV (between 0.4 and 0.6). The larger the motion amplitude, the more accurate this ITV was, suggesting it would be beneficial to adapt margins to the real motion amplitude if such a strategy is chosen. Conversely, the ITV defined as the GTV delineated on the motion averaged image was found to be much smaller than the true ITV and the larger the motion amplitude, the less accurate it was, with opposite results relative to what was observed for the first strategy (very high PPV and low SE). This approach is therefore ill-adapted for the estimation of the real tumor target volume during the respiratory cycle, confirming recent conclusions derived from observations on clinical datasets [Hanna 2012] and in contradiction to previous findings [Caldwell 2003]. This can be explained by the blurring effect of respiratory motion; as the tumor spends less time at the extreme positions during breathing, the signal is significantly weaker at these positions, resulting in smaller volumes when delineating them on the motion averaged image.

ITVs defined as the union of GTVs delineated on each respiration binned frame were the most accurate ones, in accordance with previous observations on clinical images when comparing 3D vs. 4D PET volumes [Aristophanous 2012]. Moreover, their accuracy was found to be independent of the motion amplitude. Using smaller voxels in the reconstruction grid (upsampled LR images) led to systematically worst results in all cases. This can be explained by the higher level of noise in such reconstructions, since the statistics are divided by a factor of about 8 (going from  $64\text{mm}^3$  to  $8\text{mm}^3$  voxels). This has a strong impact on the image quality, especially since individual respiratory synchronized frames are characterized by low statistical quality. Results were further significantly improved by considering motion corrected images over simple respiration synchronized frames. The union of GTVs delineated on corrected images were always closer to the true ITV, irrespective of motion amplitude, regularity or tumor to background contrast. Although both considered motion correction methods led to good results with both SE and PPV around or above 0.9 in most cases, there was a small but systematic advantage of the SR-correction over the deformation model based approach. Thus, defining the tumor target volume as the union of volumes delineated on corrected binned images seems the most accurate way of incorporating temporal information in the ITV delineation step for treatment planning.

Our study provided a full comparison of different strategies for the incorporation of respiratory motion in the PET-based ITV delineation step in RTP. This study was carried out within a simulated framework only, although as realistic as possible, and validation on even more complex cases and clinical images should now be performed. **It should be emphasized that some of the statistical tests had to be performed on very small samples ( $n=4$  for instance) and/or potentially correlated variables. These results should therefore be interpreted with caution.** Another limitation of our study is that we only investigated the delineation of PET images without considering associated anatomical/morphological datasets. One should keep in mind that radiotherapy delivery should also take into account anatomical information regarding the delineation of tumor volumes in the planning step, and actual respiratory motion monitoring during the treatment delivery step. Finally, the actual

dosimetry impact of this increased accuracy, especially on OARs sparing and the definition of tumor sub-volumes for dose painting, redistribution or boosting scenarios [Perrin 2010, Aristophanous 2011] requires further investigation that will form the basis of future studies.

### **5. Conclusion**

Different strategies of incorporating respiratory motion information during the functional internal target volume delineation on PET images for radiotherapy treatment planning were compared within a realistic simulated framework. Our results suggest that the functional tumor internal target volume defined as the union of functional gross tumor volumes delineated on motion-corrected binned images is the closest to the real functional internal target volume with respiratory motion, independently of motion regularity, motion amplitude, or tumor to background contrast.

## References

- [Aristophanous 2011] M. Aristophanous, J.T. Yap, J.H. Killoran, et al, “Four dimensional positron emission tomography: implications for dose painting of high-uptake regions”, *Int J Radiat Oncol Biol Phys* 2011;80(3):900-908.
- [Aristophanous 2012] M. Aristophanous, R.I. Berbeco, J.H. Killoran, et al, “Clinical utility of 4D FDG-PET/CT scans in radiation treatment planning”, *Int J Radiat Oncol Biol Phys* 2012;82(1):99-105.
- [Ashamalla 2005] H. Ashamalla, S. Rafla, K. Parikh, et al, “The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer”, *Int J Radiat Oncol Biol Phys* 2005;63(4):1016-23
- [Boucher 2004] L. Boucher, S. Rodrigue, R. Lecomte, F. Benard, “Respiratory gating for 3-dimensional PET of the thorax: feasibility and initial results” *J Nucl Med* 2004;45(2):214-219.
- [Bradley 2004] J. Bradley, W. L. Thorstad, S. Mutic, et al., “Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer,” *Int J Radiat Oncol Biol Phys* 2004;59(1):78-86.
- [Caldwell 2003] C.B. Caldwell, K. Mah, M. Skinner, C.E. Danjoux, “Can PET provide the 3D extent of tumor motion for individualized internal target volumes? A phantom study of the limitations of CT and the promise of PET.” *Int J Radiat Oncol Biol Phys* 2003;55(5):1381-93.
- [Chang 2009] G. Chang, T. Pan, F. Qiao, J. W. Clark, O. R. Mawlawi, “Comparison between two super-resolution implementations in PET imaging,” *Med. Phys* 2009;36(4):1370
- [de Bari 2011] B. De Bari, N. Sellal, F. Mornex, “4D-CT scan and radiotherapy for hepatocellular carcinoma: role in the definition of internal target volume (ITV)”, *Cancer Radiother.* 2011;15(1):43-8.
- [Descourt 2006] P. Descourt, W.P. Segars, F. Lamare, et al, “RTNCAT (Real Time NCAT): Implementing Real Time physiological movement of voxelized phantoms in GATE”, in *proc. IEEE NSS-MIC* 2006.
- [Green 1990] P. J. Green, “Bayesian reconstructions from emission tomography data using a modified EM algorithm,” *IEEE Trans. Med. Imag.* 1990;9(1):84–93.
- [Hanna 2012] G.G. Hanna, J.R. van Sörnsen de Koste, M.R. Dahele, et al, “Defining Target Volumes for Stereotactic Ablative Radiotherapy of Early-stage Lung Tumours: A Comparison of Three-dimensional (18)F-fluorodeoxyglucose Positron Emission Tomography and Four-dimensional Computed Tomography.” *Clin Oncol (R Coll Radiol)*. 2012 Aug;24(6):71-80
- [Hatt 2009] M. Hatt, A. Turzo, C. Roux, D. Visvikis, “A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET”, *IEEE Trans Med Imaging* 2009;28(6):881-893.
- [Hatt 2010] M. Hatt, C. Cheze le Rest, P. Descourt, et al, “Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications,” *Int J Radiat Oncol Biol Phys* 2010;77(1):301-308.
- [Hatt 2011a] M. Hatt, C. Cheze-Le Rest, N. Albarghach, O. Pradier, D. Visvikis “PET functional volume delineation: a robustness and repeatability study”, *Eur J. Nucl. Med. Mol. Imaging* 2011;38(4):663-672.

- [Hatt 2011b] M. Hatt, C. Cheze le Rest, A. van Baardwijk, P. Lambin, O. Pradier, D. Visvikis, "Impact of tumor size and tracer uptake heterogeneity in 18F-FDG PET and CT Non-Small Cell Lung Cancer tumor delineation", *J Nucl Med* 2011;52(11):1690-7.
- [Huber 1981] P. J. Huber, "Robust Statistics," John Wiley & Sons, New York, 1981.
- [Jan 2004] S. Jan, G. Santin, D. Strul, et al, "GATE: a simulation toolkit for PET and SPECT" *Phys Med Biol* 2004;49:4543-4561.
- [Jarrit 2006] P.H. Jarrit, K. Carson, D. Visvikis, "The role of PET/CT scanning in radiotherapy treatment planning", *Br J Rad* 2006;79:S27-S35.
- [Kang 2003] M. G. Kang and S. Chaudhuri, "Super-resolution image reconstruction," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 19-20, 2003.
- [Keall 2006] P. J. Keal, G. S. Mageras, J. M. Balter, et al., "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med Phys* 2006;33(10):3874-3900.
- [Lamare 2006] F. Lamare, A. Turzo, Y. Bizais, C. Cheze Le Rest, D. Visvikis, "Validation of a Monte Carlo simulation of the Philips Allegro/GEMINI PET systems using GATE," *Phys Med Biol* 2006;51(4):943-962.
- [Lamare 2007a] F. Lamare, T. Cresson, J. Savean, et al, "Respiratory motion correction for PET oncology applications using affine transformation of list mode data." *Phys Med Biol* 2007;52(1):121-40.
- [Lamare 2007b] F. Lamare et al., "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," *Phys Med Biol* 2007;52(17):5187-5204.
- [Lamb 2011] J.M. Lamb, C. Robinson, J. Bradle, et al, "Generating lung tumor internal target volumes from 4D-PET maximum intensity projections" *Med Phys* 2011;38(10):5732-7.
- [Le Maitre 2009] A. Le Maitre, W.P. Segars, S. Marache, et al, "Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications", *Proceedings of the IEEE* 2009;97(12):2026-2038.
- [Nehmeh 2004] S.A. Nehmeh, Y.E. Erdi, T. Pan, et al, "Four-dimensional (4D) PET/CT imaging of the thorax", *Med Phys*. 2004;31(12):3179-86.
- [Park 2008] S.J. Park, D. Ionascu, J. Killoran, et al, "Evaluation of the combined effects of target size, respiratory motion and background activity on 3D and 4D PET/CT images." *Phys Med Biol* 2008;53(13):3661-79.
- [Perrin 2010] R. Perrin, P.M. Evans, S. Webb, M. Partridge, "The use of PET images for radiotherapy treatment planning: an error analysis using radiobiological endpoints" *Med Phys* 2010;37(2):516-31.
- [Qiao 2006] F. Qiao, T. Pan, W.J. Clark Jr, O.R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies", *Phys. Med. Biol.* 2006;51:3769-3783.
- [Riegel 2010] A.C. Riegel, M.K. Bucci, O.R. Mawlawi, et al, "Target definition of moving lung tumors in positron emission tomography: correlation of optimal activity concentration thresholds with object size, motion extent, and source-to-background ratio", *Med Phys*. 2010;37(4):1742-52.
- [Segars 2009] W.P. Segars, D.S. Lalush, E.C. Frey, et al, "Improved Dynamic Cardiac Phantom Based on 4D NURBS and Tagged MRI." *IEEE Trans Nucl Sci.* 2009;56(5):2728-2738.



- [Steenbakkers 2006] R.J.H.M. Steenbakkers, J.C. Duppen, I. Fitton, et al, “Reduction of observer variation using matched CT-PET for lung cancer delineation: a three dimensional analysis,” *Int J Radiat Oncol Biol Phys* 2006;64(2):435-448.
- [van Baardwijk 2007] A. van Baardwijk, G. Bosmans, L. Boersma, et al., “PET-CT–Based Auto-Contouring in Non–Small-Cell Lung Cancer Correlates With Pathology and Reduces Interobserver Variability in the Delineation of the Primary Tumour and Involved Nodal Volumes,” *Int J Radiat Oncol Biol Phys* 2007;68(3):771-778.
- [van Loon 2011] J. van Loon, A. van Baardwijk, L. Boersma, M. Ollers, P. Lambin, D. De Ruyscher, “Therapeutic implications of molecular imaging with PET in the combined modality treatment of lung cancer.”, *Cancer Treat Rev.* 2011;37(5):331-43.
- [Walter 1999] K. Walter, S. Schmidt, N. Licht, C. Nieder, B. Motaref, D. Hellwig, et al., “<sup>18</sup>F-deoxyglucose positron emission tomography (FDG-PET) for the planning of radiotherapy in lung cancer: high impact in patients with atelectasis,” *Int J Radiat Oncol Biol Phys* 1999;44(3):593-597.
- [Wallach 2010] D. Wallach, F. Lamare, C. Roux, D. Visvikis, “Comparison between reconstruction-incorporated super-resolution and super-resolution as a post-processing step for motion correction in PET” , in *proc. IEEE NSS/MIC 2010*:2294-2297.
- [Wallach 2012] D. Wallach, F. Lamare, G. Kontaxakis, D. Visvikis, “Super-resolution in respiratory synchronized positron emission tomography” *IEEE Trans Med Imaging* 2012;31(2):438-48.
- [Wolthaus 2005] J.W. Wolthaus, M. van Herk, S.H. Muller, et al. “Fusion of respiration-correlated PET and CT scans: correlated lung tumour motion in anatomical and functional scans”, *Phys Med Biol* 2005;50(7):1569-83.